DOCUMENT RESUME

ED 409 335 TM 026 792

AUTHOR Foegen, Anne; Deno, Stanley L.
TITLE General Outcome Measures in Secondary Mathematics: An Exploration of Relations to the NCTM Goals.
PUB DATE Mar 97
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Objectives; Junior High Schools; *Mathematics; *Measurement Techniques; Middle Schools; *Outcomes of Education; *Performance Based Assessment; Problem Solving; Student Attitudes; Teacher Attitudes; Teacher Expectations of Students; Thinking Skills; Urban Schools
IDENTIFIERS *Middle School Students; *National Council of Teachers of Mathematics

ABSTRACT

As part of a larger study exploring the degree to which the general outcome measurement paradigm might be extended to address assessment issues in secondary mathematics, this study involved determining the degree to which simple performance measures would serve as indicators of general proficiency in mathematics and success in achieving the goals outlined by the National Council of Teachers of Mathematics (NCTM). These goals state that students should: (1) learn to value mathematics; (2) become confident in their ability to do mathematics; (3) become mathematical problem solvers; (4) learn to communicate mathematically; and (5) learn to reason mathematically. Participants were 100 urban middle school students. Four general outcome measures were used, including a basic mathematics operations task, a basic estimation task, modified estimation tasks, and teacher ratings. Results of the analyses support the conclusion that low to moderate relations exist between students' scores on the general outcome measures and their success in achieving the NCTM goals. The strongest relations exist between performance on the basic facts and estimation tasks and the teacher's perceptions of their proficiency and problem-solving status. General outcome measures like the ones studied are sufficiently related to teachers' judgments of student status in mathematics that they could use to monitor progress toward the NCTM goals. Although the general outcome measures do not meet all the criteria for new assessment tools proposed by the NCTM, the tasks explored in this study allow for more frequent and repeated administration. (Contains 8 tables and 54 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 409 335

Running head: GENERAL OUTCOME MEASURES AND NCTM GOALS

General Outcome Measures in Secondary Mathematics: An Exploration of Relations to the NCTM Goals

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Anne Foegen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Anne Foegen

Iowa State University

Stanley L. Deno

University of Minnesota

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper/poster presented at the Annual Meeting of the American Educational Research Association, March, 1997. Chicago, IL.

Correspondence concerning this paper should be addressed to Anne Foegen, Department of Curriculum and Instruction, Iowa State University, N134 Lagomarcino Hall, Ames, IA, 50011.

TM026792

General Outcome Measures in Secondary Mathematics: An Exploration of Relations to the NCTM Goals

Recent years have seen increased attention to outcomes produced by the public education system. Improved student achievement has served as a focal point for national educational goals (Riley, 1994; U.S. Department of Education, 1991; 1993) and legislative appropriations (Schnaiberg, 1994, September 28). The demand for increased accountability and improved outcomes has not been exclusive to the field of general education. Special educators also find themselves urged to demonstrate the effects of their programs and practices (Ysseldyke, Thurlow, & Shriner, 1992).

These calls for improved outcomes are particularly pertinent given the findings of research on adult outcomes for students with disabilities. The literature on transition to postschool environments has identified several alarming trends regarding the difficulties students with disabilities encounter in attempting to achieve adult outcomes similar to those of their nondisabled peers. Students with disabilities are more likely than their nondisabled peers to drop out of school, be employed in unskilled or semiskilled jobs, be employed on a part-time basis, and live at home with their parents (deBettencourt, Zigmond, & Thornton, 1989; Sitlington, Frank, & Carson, 1993, Wagner, 1990). Ethnographic studies examining the experiences of particular individuals in greater detail have identified similar findings, with many of the students with disabilities involved in transient work and post-secondary schooling experiences, and dependent on parents or relatives for food and housing (Lichtenstein, 1993; Zetlin & Hosseini, 1989). Heal and Rusch (1995) found that students' personal characteristics (including academic skills) were highly influential in predicting the postschool employment of students with disabilities. Recent studies (Blackorby & Wagner, 1996) have shown some increases for students with disabilities in the areas of wages, employment rates, postsecondary schooling and residential independence, but significant gaps continue to exist in levels of educational attainment. If outcomes for secondary students with disabilities are to be improved, teachers will need better methods of monitoring student progress and identifying students in need of additional intervention.

In the field of mathematics education, concerns about current levels of student achievement and the curricula and instructional practices existing in many schools led to the publication of standards for curriculum and evaluation, teaching, and assessment by the National Council of Teachers of Mathematics (NCTM, 1989, 1991, 1995). These documents have outlined a vision for mathematics education that moves away from the computation-laden curriculum of the past and toward a challenging, concept-driven curriculum that enables students to solve problems and reason logically. The standards "articulate five general goals for all students: (1) that they learn to value mathematics, (2) that they become confident in their ability to do mathematics, (3) that they become mathematical problem solvers, (4) that they learn to communicate mathematically, and (5) that they learn to reason mathematically" (NCTM, 1989, p. 5). Although the standards themselves have not been accepted without criticism (Carnine, 1992; Hofmeister, 1993; Shriner, Kim, Thurlow, & Ysseldyke, 1992, 1993), the difficulties with mathematics instruction that they were developed in response to are rarely questioned.

The emphasis on revising mathematics instruction has been accompanied by a push toward "authentic" assessment strategies involving open-ended tasks, checklists, interviews, extended investigations, and portfolios (Cain & Kenney, 1992; Clarke, 1992; NCTM, 1995). Spurred in part by dissatisfaction with multiple choice standardized tests that assess computation and isolated skills, proponents of reform have advocated for procedures that are based on student performance, more closely tied to the curriculum, representative of realistic mathematical tasks, and more useful to teachers for improving instruction (Kulm, 1994; Lesh & Lamon, 1992; Romberg, 1992). While descriptive writings outlining the proposals are abundant, there are few examples in which the proposals are operationalized into assessment procedures with reported levels of reliability and validity. Studies of the technical adequacy of performance-based assessment, however, have illustrated tremendous difficulties in developing reliable and valid procedures (Baxter, Shavelson, Herman, Brown, & Valadez, 1993).

General outcome measurement (Fuchs & Deno, 1991) has been developed in the field of special education as a means of assessing the degree to which students are achieving the broader,

generalized outcomes toward which a curriculum is directed. Curriculum-Based Measurement (CBM; Deno, 1985) represents one form of general outcome measurement that has an extensive research base (Shinn, 1989). By collecting data drawn from brief samples of student work, teachers are able to examine the progress of individual students and make timely changes in instruction to effect enhanced student progress in reading (Fuchs, Fuchs, Hamlett, & Ferguson, 1992), mathematics (Fuchs, Fuchs, Hamlett, & Stecker, 1991), written expression (Tindal & Parker, 1989a) and spelling (Fuchs, Fuchs, Hamlett, & Allinder, 1991). Although empirical support for CBM at the elementary level is extensive, only a small number of studies have examined extensions of general outcome measurement to the secondary level.

Researchers investigating the application of general outcome measurement at the secondary level have explored means of evaluating student performance in the content areas of English, science, and social studies. The measures developed for indexing students' academic success have included reading aloud and classroom study tasks (Espin, 1990, Espin & Deno, 1993); oral reading, maze, and vocabulary (Espin & Foegen, 1996); written retells (Tindal & Parker, 1989b); and essays, chapter tests, and perception probes (Nolet & Tindal, 1994; Tindal & Nolet, 1995). No general outcome measurement studies of have been conducted at the secondary level in mathematics.

The research described here was one component of a larger study exploring the degree to which the general outcome measurement paradigm might be extended to address assessment issues in secondary mathematics (Foegen, 1995). The present study involved determining the degree to which simple performance measures would serve as indicators of general proficiency in mathematics and success in achieving the goals outlined by the NCTM. The performance measures developed for this study assessed students' fluency with estimation and basic facts. Estimation was selected as the primary focus of the study because it is representative of the type of mathematical skill that is widely applied by adults in daily living situations. B. Reys (1992) noted that "estimation is a basic skill, and its growing importance in a technological society is recognized. It is used much more than exact computation" (p. 281). In fact, B. Reys (1992) estimated that

more than 80% of all mathematical applications use estimation rather than exact computation. Fluency with facts was identified in order to examine the utility of a measure drawn from CBM at the elementary level and because research in cognitive science has supported the importance of a certain level of automaticity in basic skills to attaining higher levels of performance (Gagne, 1983; Glaser, 1987; Resnick, 1989). Studies in reading (LaBerge & Samuels, 1974) have suggested that automatic decoding skills allow students to focus increased attention on the task of comprehension. Similarly, in mathematics, the "ability to succeed in higher-order skills appears to be directly related to the efficiency at which lower-order processes are executed" (Hasselbring, Goin, & Bransford, 1988, p. 1). Applying the information processing approach drawn from cognitive psychology, it was hypothesized that students who fluently computed basic facts would be more likely to have greater information processing capacity available for addressing more complex problem solving tasks.

Four general outcome measures were examined in the study: a basic facts task and three forms of an estimation task that included both computation and problem solving. Two research questions served as the foundation for the study. First, what is the relation between the data produced by the general outcome measures and students' achievement of the NCTM goals? Second, what is the relative efficacy of the general outcome measures, either separate or combined, in predicting proficiency in mathematics and success in attaining the NCTM goals?

Method

Participants and Setting

Participants for the study were students in the classes of a middle school mathematics teacher. The school in which the study took place was located in an urban district in a large metropolitan area. The school's population, which numbered approximately 1,000 students, was diverse in racial and ethnic composition. Data collected from the district's records was used to index the socioeconomic status of the school and revealed that approximately three-quarters of the students qualified for a "free and reduced lunch" program. Of the total school population (not

including a district-wide setting for students with severe behavioral disorders housed in the building) 9.9% of the students were receiving special education services.

Of the 128 students in the teacher's classes who were eligible to participate in the study, the 100 students for whom permission had been obtained comprised the sample for this study. The sample included 48 males and 52 females. Forty-two percent of the students in the sample were of European American descent, 39% of African American descent, 12% of Asian American descent, 4% of American Indian descent, and 3% of Hispanic American descent. The mean age of the participants was 13.52 years, with a range of 11.7 to 15.4 years. The sample comprised primarily 7th and 8th grade students, with 9, 49, and 41 students in the 6th, 7th, and 8th grades, respectively. Twelve students were receiving special education services. These students had been identified as having learning disabilities or behavioral disorders using state guidelines.

The four classes taking part in the study included one algebra class, one pre-algebra class, and two general math classes. The teacher and his classes were currently participating in a larger project (Robinson & Deno, 1992, November) exploring the use of a group-based communication technology system to support students with mild disabilities in general education classrooms. The technology system (Discourse®, 1990) had been in place in the teacher's classroom since the onset of the school year and was used daily for instructional activities. Although the data for this research were collected through the technology system, the system itself was not a primary focus of the study. All students had used the technology extensively and were familiar with its operation.

Materials

General Outcome Measures

Four general outcome measures were used in the study. Foegen (1995) examined the reliability and criterion validity of these measures. When scores were aggregated across two testing events, test-retest and alternate forms reliability estimates were in the .79 to .88 range. Criterion validity estimates, also based on the aggregation of two scores, ranged from .44 to .63 when the scores were correlated with scores from the Mathematical Computation and Mathematical

Concepts and Applications subtests of the California Achievement Test. Each of the general outcome measures is described in detail below. Critical features of each task and sample problems are presented in Table 1. A single measure tapped students' fluency with basic facts, while three different measures were used to examine students' estimation skills.

Insert Table 1 about here

Basic Math Operations Task (BMOT). The BMOT was designed to indicate students' accuracy and fluency in mental computation of whole number facts (0-9) in each of the four operations. Two parallel forms of the BMOT were developed for use in the study by randomly selecting 20 single digit combinations for each of the operations of addition, subtraction, multiplication, and division. The 80 items were randomly ordered for placement on the probe. Students had one minute to respond to each BMOT. The probes utilized a constructed response format, in which students were presented with the problems and responded with numerical answers.

Basic Estimation Task (BET). The BET was designed to assess students' accuracy and fluency in using estimation, mental computation, and mathematical judgment to select responses to computation or problem solving problems. A single probe, drawn from a series of probes developed for the larger study, was administered during each testing session. The BET problems crossed all four operations, but were limited to whole numbers. A range of difficulty was represented, with some of the problems involving two-, three-, and four-digit numbers. Computation problems were created using the random number generator function on a handheld calculator. Word problems were developed around three general topics: money and purchasing; sports, hobbies, and music; and school and work. Three response options were presented, varying in magnitude to facilitate students' use of estimation or "number sense" (Sowder, 1992) to quickly select the correct answer. The correct answer was the result that would be obtained if the exact computation was carried out. Distracters were devised to represent likely student errors

(e.g., faulty algorithm, incorrect operation, digits out of order). Students had three minutes to respond to 40 items by selecting the letter (a/b/c) corresponding to the correct answer.

Modified Estimation Tasks (METs). The METs (Form A and Form B) represented variations of the BET. The major differences included an expanded range of number forms (rational in addition to whole numbers), a higher proportion of word problems, and the exclusion of basic facts problems. The problems for each form of the MET were identical. They were developed by selecting numbers that could be rounded to obtain a problem that could be computed using mental arithmetic, thus encouraging students to use reformulation and other estimation strategies (R. Reys, Rybolt, Bestgen, & Wyatt, 1982). The METs differed from each other only in response format. The MET-A provided alternatives in the form of exact numbers (e.g., 16, 247, 3562), paralleling the format used in the BET and constructed using similar procedures. MET-B response alternatives were rounded numbers that differed from each other by a power of ten (e.g., 40, 400, 4,000) and represented a response format often used in tests of estimation (R. Reys, 1986; Schoen, Blume, & Hoover, 1990). Two parallel forms of each type of task (MET-A and MET-B) were developed. Students had 3 minutes to respond to 40 problems by selecting the letter corresponding to the correct answer.

Teacher Ratings

Data on mathematics proficiency and students' success in achieving the NCTM curricular goals were obtained from teacher ratings. For each of the students, the teacher used a Likert scale to rate (1=very low, 5=very high) the student's 1) overall proficiency in mathematics, 2) value for mathematics, 3) confidence in his/her mathematics ability, 4) mathematical problem solving ability, 5) mathematical communication ability, 6) mathematical reasoning ability. The last five ratings corresponded to the goals advocated by the NCTM. Rating data were collected from the teacher during the same week in which the general outcome measures were administered.

Procedures

On two occasions during one week in the spring of 1995, students completed a series of measurement tasks. The tasks included the BET, two parallel forms of the BMOT, and two parallel

forms of either the MET-A or the MET-B. Each of the four classes was randomly assigned to one of the two MET forms. The task sequence began with the BET for all four classes, with the order of the BMOT and MET counterbalanced across classes. The same sequence of administration was used in each testing session. Prior to the first administration of each of the unfamiliar tasks (BMOT and MET), students were provided with a brief (5 minute) instructional session during which they were shown examples of the types of problems in the task, provided a demonstration of the solution process, and allowed to practice responding to sample items.

Data for the study were collected via a group response technology system (Discourse®, 1990; see Foegen & Hargrave, 1997; Robinson, 1991, 1994) being used for instructional purposes in the teacher's classroom as part of the larger study. The group response technology used networking to link individual student terminals to a teacher work station. Each terminal consisted of an eight line display (40 characters per line) and a standard keyboard. Students viewed each task on the display and were able to control the pacing of item presentation by moving from one screen to the next using simple keyboard commands. The system permitted the investigator to precisely control the timing of the tasks and allowed the students' responses to be immediately recorded, scored, and saved to disk. Foegen (1995) examined the scoring accuracy of the computer technology and obtained interscorer (computer vs. human) reliability estimates ranging from 99.3 to 99.9%.

Analyses

The first research question for the study addressed the degree to which students' scores on the general outcome measures related to teacher ratings of their overall proficiency in mathematics and their achievement of the curricular goals for mathematics identified by the NCTM. The second question explored the efficacy of using the measures singly, or in combination, to predict the teacher ratings. The primary analyses involved correlation and multiple regression. Correlation coefficients were used to explore relations between students' scores on the general outcome measures and the teacher ratings. Multiple regression analyses permitted an examination of the

degree to which the general outcome measures could be used singly or in combination to predict teacher ratings.

Results

The means and standard deviations for scores obtained from the general outcome measures and the teacher ratings are presented in Table 2. Mean scores on the BET reflected incremental increases across grades; the same was not true of the other measures. One factor influencing this result may have been students' familiarity with the BET. As part of the larger study, students had completed one BET per week for the 24 weeks of school preceding the data collection period for this study. In contrast, the BMOT and METs were new tasks to which the students had not previously been exposed. The limited variability in the scores obtained on the teacher ratings should be noted when evaluating the results of subsequent analyses. This result is a particular problem for rating scales, which tend to be cross sectional, rather than longitudinal, in character.

Insert Table 2 about here

The focus of the first research question was on the degree to which students' scores on the general outcome measures related to their general proficiency in mathematics and their success in achieving the NCTM goals, as judged by their mathematics teacher. Pearson product-moment correlation coefficients between the general outcome scores and the teacher ratings are presented in Table 3. Reliable low to moderate relations were identified between students' scores on the four general outcome tasks and the teacher's ratings of students' mathematical Proficiency, Communication, Confidence, Problem-Solving, and Reasoning. One exception was the "Valuing Mathematics" rating, for which none of the relations were reliable. We subsequently excluded this scale from further analyses.

Insert Table 3 about here

The results of the correlational analyses indicated that the general outcome measures were very similar in the degree to which they predicted ratings of students' achievement of the NCTM goals. With the exception of the MET-A, for which the coefficients were somewhat lower, the BMOT, BET and MET-B appeared to predict teacher ratings with comparable levels of accuracy. Differences were observed in the six scales on which the teacher rated student performance. The general outcome measures were more strongly related to the Proficiency, Problem-Solving, and Reasoning scales than they were to the Communication, Confidence, and Valuing Mathematics scales.

The relative efficacy of using the general outcome measures, separately or in combination, to predict teacher ratings was examined using a series of regression analyses. Our purpose was to determine whether any of the general outcome measures contributed uniquely to the teacher's perceptions of student status on the NCTM goals. Because each class was randomly assigned to one of the two Modified Estimation Tasks (either MET-A or MET-B), it was not possible to include both of these measures in the same regression analysis. Rather than double the number of analyses conducted, we decided to drop the MET-A from further consideration. This decision was based on the results of the correlational analyses (where MET-A correlations with teacher ratings were much lower than those obtained for MET-B) and technical adequacy data on the measures obtained by Foegen (1995) indicating that the MET-B was a more reliable and valid measure than the MET-A.

The first set of regression analyses, presented in Table 4, examined the contribution of each of the three general outcome measures to predicting the five teacher ratings. No constraints were applied to the equations, therefore the order in which the general outcome measures were entered into the regression equation reflected the relative strength with each measure related to the teacher rating. In each case, either the BMOT or the BET entered into the equation first. The MET-B was always the last variable entered, indicating it accounted for the smallest amount degree of unique variance in the teacher rating. With the exception of Problem-Solving, all of the teacher ratings were effectively predicted by a single general outcome measure; the addition of subsequent

measures did not contribute to the predictive power of the equation at a level that was statistically significant. In the case of the Problem-Solving ratings, the predictive power of the BET and BMOT scores in combination was significantly higher than that of the BET alone. The Confidence ratings differed noticeably from the others in that a very small proportion (14%) of the variance in confidence ratings was accounted for by the general outcome measures. For the remaining scales, however, the general outcome measures accounted for 30 to 50% of the variability in the ratings.

Insert Table 4 about here

To explore the degree of overlap associated with the general outcome measures, we next conducted a series of analyses in which we controlled the order in which the general outcome measures were entered into the regression equation. Because of the increased potential for error associated with a large number of analyses, we dropped the Confidence ratings from this portion of the analysis. For each of the remaining teacher ratings (Proficiency, Communication, Problem-Solving, and Reasoning), we ran three regression analyses in which one of the general outcome measures was forced into the equation first, with the remaining two measures allowed to vary in order of entry. The results of these analyses, presented in Tables 5 though 8, indicate the degree to which the unique variance accounted for by the measures overlaps and the usefulness of using the measures in combination to enhance predictive ability.

Insert Tables 5, 6, 7, and 8 about here

The first series of regression analyses (Table 5) used the Proficiency rating as the dependent variable. The results indicated that the BMOT alone was able to account for a majority of the variance, with no additional predictive power gained by adding the remaining measures. When either the BET or the MET-B was entered into the equation first, the BMOT accounted for a

statistically significant proportion of unique variance when added to the equation. The increases obtained by adding the BMOT ranged from .09 for the BET to .17 for the MET-B.

In the second set of analyses (Table 6), the Communication rating served as the dependent variable. The pattern of results differed slightly than those obtained in the first set of analyses. For the Communication ratings, the BET was the single strongest predictor, with no statistically significant increases in variance accounted for by adding the remaining measures. However, when either the BMOT or the MET-B were entered first, the BET added to the equation. Increases in the proportion of variance accounted for were .08 and .17 for the BMOT and MET-B, respectively.

The third set of analyses (Table 7) used the Problem-Solving rating as the dependent variable. In this case, the BMOT and BET alone accounted for comparable proportions of the variance in Problem-Solving ratings, and each explained additional unique variance when added to the other. The increases in proportion of explained variance were in the .07 to .08 range. When the MET-B was entered first, both the remaining measures added to the equation, with the proportion of variance accounted for increasing by .22 and .07 for the BET and BMOT, respectively.

The final set of analyses (Table 8) examined the contribution of the general outcome measures to the prediction of Reasoning ratings. The pattern of results was identical to that obtained for the Proficiency ratings. The BMOT was the single best predictor, with no statistically significant improvements in prediction obtained by adding the BET or MET-B. When either the BET or MET-B was entered first, the BMOT accounted for additional unique variance when added to the equation. Increases in the proportion of variance accounted for ranged from .11 when the BMOT was added to the BET to .23 when it was added to the MET-B.

Taken together, the results of these analyses indicate that there was a high degree of overlap in the variance accounted for by the general outcome measures. For three of the ratings, the variance accounted for by the single strongest measure was not increased to a statistically significant degree by adding either of the remaining measures. Furthermore, when this strongest measure was <u>not</u> the first entered into the equation, increases in <u>total</u> variance achieved by adding it

to the equation were minimal, ranging from .01 to .04. The exception to this pattern was Problem-Solving, which was best predicted by a combination of the BMOT and the BET. The failure of the MET-B to contribute to the prediction of any of the teacher ratings was surprising, given the strength of the correlations reported in Table 3. The regression results suggest that the variance in teacher ratings explained by the MET-B overlapped entirely with that explained by the BMOT and the BET.

The results of these analyses are useful when examining the efficacy of using the measures in combination. Because of the high degree of overlap present in the prediction of Proficiency, Communication, and Reasoning, it appears that teacher ratings of these three areas are most efficiently predicted by single general outcome measures. The BMOT represented the single best measure for predicting Proficiency and Reasoning, while Communication was best predicted by the BET. For the area of Problem-Solving, the fact that the predictive ability of either the BMOT or the BET was improved by adding the other measure indicates there was less overlap in the variance accounted for by these two general outcome measures.

Discussion

This study explored relations between four general outcome measures and teacher ratings of student success in attaining the NCTM's goals for mathematics, which state that students should 1) learn to value mathematics, 2) become confident in their ability to do mathematics, 3) become mathematical problem solvers, 4) learn to communicate mathematically, and 5) learn to reason mathematically (NCTM, 1989). Two research questions were addressed in the study. First, what is the relation between the data produced by the general outcome measures and students' proficiency in mathematics and achievement of the NCTM goals? Second, what is the relative efficacy of the general outcome measures, either separate or combined, in predicting proficiency in mathematics and status on the NCTM goals?

The results of the analyses support the conclusion that low to moderate relations exist between students' scores on the general outcome measures and their success in achieving the NCTM goals. The strongest relations apparently exist between students' performance on the basic

facts and estimation tasks and the teacher's perceptions of their Proficiency and Problem-Solving status. The teacher's perceptions of students' Value for Mathematics seem to be unrelated to their performance on the general outcome measures. The relations for Confidence were considerably lower than those obtained for perceptions of Proficiency, Communication, Problem-Solving, and Reasoning. Apparently, the affective NCTM goals (Confidence and Value for Mathematics) are somewhat less a function of achievement than might be expected.

When we examine the question of how well the general outcome measures can be used in combination to predict student status on the NCTM goals, we are left to conclude that the apparent differences in mathematics skills tapped by the types of test items used do not translate into empirical differences in making predictions. Single measures can predict teacher assessment of mathematics competence about as well as combinations of measures. For example, a simple measure of fluency with basic facts predicts ratings of overall Proficiency and Reasoning, while estimation performance best predicts ratings of mathematical Communication. In contrast, a combination of facts and estimation may be the most successful means of predicting ratings of Problem-Solving. These results provide additional support for those who argue that accurate and fluent mental manipulation of basic math operations is fundamental to success in mathematics class performance and is an aspect of developing mathematical competence that should not be ignored.

When interpreting the present results, it is useful to consider previous findings relating student performance data to other indicators of mathematics achievement. Although no research was identified that directly examined relations between measures of student performance and achievement of the NCTM goals, studies of curriculum-based student performance measures used at the elementary level can provide a context for evaluating the results obtained here. The present findings are comparable to results obtained in previous studies of the criterion validity of computation-based CBM measures at the elementary level. Marston (1989) reported mid-range correlations between computation-based CBM measures and standardized test scores. Foegen (1995) obtained similar results in an investigation of the criterion validity of facts and estimation general outcome measures at the middle school level. The comparability of the present findings are

notable, given the limited range of scores for the ratings and the estimation measures, which likely suppressed the obtained correlations.

The results of this research must be considered in light of several limitations associated with the study. First, relatively few samples of student performance were obtained. This is particularly relevant for results obtained for the modified estimation tasks, which were unfamiliar to students. Additional practice with these tasks may influence the scores obtained. The second limitation involves the teacher ratings, which were based on a single teacher's evaluations of each of his students on six rating scales. Although the use of a single teacher increased the consistency with which the ratings were applied, the results may be biased by this teacher's pattern of responding. Examination of the data indicated the teacher tended to rate individual students similarly across all six scales. Future research should explore alternative means of assessing the degree to which students are attaining the goals identified by the NCTM. At this point in time, few alternatives exist for assessing these outcomes. A third limitation is the restricted range of scores obtained for the teacher ratings (which ranged from 1 to 5) and some of the general outcome measures. For the modified estimation tests, in particular, the effective range was from zero to ten responses for two-thirds of the students. Floor effects for these measures should be considered in interpreting the present results; options for increasing the range of scores should be explored in future research. Finally, the use of the technology for data collection may represent a possible confounding factor in the study. Although the students were familiar with the operation of the system, the degree to which scores obtained via the technology are comparable to those obtained using pencil and paper methods is unknown.

The results of this study support the conclusion that general outcome measures like those investigated are sufficiently related to teachers' judgments of student status in mathematics that they could be used to monitor progress toward the NCTM goals. Data from these brief samples of student performance may be useful to teachers as they make decisions related to mathematics instruction. For example, they could be used as a screen to identify students at risk in mathematics who will need higher levels of assistance. The data might also be useful in decisions involving the

placement of students with disabilities in general education or special education settings for their mathematics education. As teachers select which of the general outcome measures to use for decision making, their decisions may be heavily influenced by the feasibility of developing, administering, and scoring the various types of measures. Although the estimation tasks are likely to have higher face validity for secondary teachers, the degree of overlap in the variance accounted for by the measures suggests that the basic facts task may represent a more efficient alternative for teachers in terms of feasibility.

Further research is needed to explore the degree to which the general outcome measures are sensitive to individual student progress. If they are, great potential exists for using the data for formative evaluation of program effectiveness for individual students, perhaps effecting the increased levels of student achievement obtained through the implementation of CBM at the elementary level. Refinement of the general outcome measures is needed, particularly the response format options used with the estimation tasks and issues related to face validity of the measures. Finally, it is important to explore the development of other external criteria (in addition to teacher ratings) that might serve as alternative means of evaluating student success in mathematics.

Although general outcome measurement does not meet all the criteria for new assessment tools proposed by the NCTM (1995), the basic facts and estimation tasks explored here offer several advantages over existing assessment options. When compared to traditional norm-referenced tests, general outcome measures offer the added advantages of allowing for more frequent and repeated administration. They can also be more directly tied to general curricular goals, and are more likely to produce data useful to teachers for instructional decision making. When compared to the performance-based alternative assessment practices proposed by the NCTM and in the mathematics education literature, general outcome measures are logistically feasible in that they require minimal commitments of time for administration and scoring; they possess known (and acceptable) levels of technical adequacy; and, finally, it is possible to develop multiple forms of the measures so that repeated measurement of growth is possible.

As the field of mathematics education moves toward the reforms necessary to accomplish the NCTM goals, new alternatives for assessment are being developed. The general outcome measures examined in this study represent one means of measuring students' performance in mathematics. Indeed, as Fuchs and Fuchs (1996) have proposed, the combination of general outcome measurement with performance assessment may provide teachers with complementary forms of assessment information from which to evaluate their students' mathematics performance and make better-informed instructional decisions.

REFERENCES

Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. Journal for Research in Mathematics Education, 24, 190-216.

Blackorby, J., & Wagner, M. (1996). Longitudinal postschool outcomes of youth with disabilities: Findings from the National Longitudinal Transition Study. Exceptional Children, 62, 399-413.

Cain, R. W., & Kenney, P. A. (1992). A joint vision for classroom assessment. Clarke, D. J. (1992). Activating assessment alternatives in mathematics. Arithmetic Teacher, 39 (6), 24-29.

Carnine, D. (1992). Expanding the notion of teachers' rights: Access to tools that work. Journal of Applied Behavior Analysis, 25, 13-19.

Clarke, D. J. (1992). Activating assessment alternatives in mathematics. Arithmetic Teacher, 39 (6), 24-29.

deBettencourt, L. U., Zigmond, N., & Thornton, H. (1989). Follow-up of postsecondary-age rural learning disabled graduates and dropouts. Exceptional Children, 56, 40-49.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Discourse [Computer groupware]. (1990). Milwaukee, WI: Discourse Technologies.

Espin, C. A. (1990). Reading aloud from text as an indicator of achievement in the content areas. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Espin, C. A., & Deno, S. L. (1993). Performance in reading from content area text as an indicator of achievement. Remedial and Special Education, 14 (6), 47-59.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. Exceptional Children, 62, 497-514.

Foegen, A. M. (1995). Reliability and validity of three general outcome measures for low-achieving students in secondary mathematics. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Foegen, A., & Hargrave, C. P. (1997). Discourse: Groupware for interactive classroom communication [Review of the computer groupware Discourse]. Journal of Computing in Teacher Education, 13 (2), 30-32.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children, 57, 488-500.

Fuchs, L. S., & Fuchs, D. (1996). Combining performance assessment and curriculum-based measurement to strengthen instructional planning. Learning Disabilities Research & Practice, 11 183-192.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1989). The reliability and validity of skills analysis within curriculum-based measurement. Diagnostique, 14, 203-221.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement using a reading maze task. Exceptional Children, 58, 436-450.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. American Educational Research Journal, 28, 617-641.

Gagne, R. M. (1983). Some issues in the psychology of mathematics instruction. Journal for Research in Mathematics Education, 14, 7-18.

Hasselbring, T. S., Goin, L. I., & Bransford, J. D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. Focus on Exceptional Children, 20 (6), 1-7.

Heal, L. W., & Rusch, F. R. (1995). Predicting employment for students who leave special educational high school programs. Exceptional Children, 61, 472-487.

Hofmeister, A. M. (1993). Elitism and reform in school mathematics. Remedial and Special Education, 14 (6), 8-13.

Kulm, G. (1994). Mathematics assessment: What works in the classroom. San Francisco: Jossey-Bass.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Lesh, R. & Lamon, S. J. (1992). Assessment of authentic performance in school mathematics. Washington, DC: AAAS Press.

Lichtenstein, S. (1993). Transition from school to adulthood: Case studies of adults with learning disabilities who dropped out of school. Exceptional Children, 59, 336-347.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), Curriculum-Based Measurement: Assessing Special Children (pp. 18-78). New York: Guilford.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standard for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). Professional standards for teaching mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). Assessment standards for school mathematics. Reston, VA: Author.

Nolet, V., & Tindal, G. (1994). Instruction and learning in middle school science classes: Implications for students with disabilities. Journal of Special Education, 28, 166-187.

Reys, B. (1992). Estimation. In T. R. Post (Ed.), Teaching mathematics in grades K-8: Research-based methods (pp. 279-301). Boston: Allyn and Bacon.

Reys, R. E. (1986). Evaluating computational estimation. Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 225-238). Reston, VA: NCTM.

Reys, R. E., Rybolt, J. F., Bestgen, B. J., & Wyatt, J. W. (1982). Processes used by good computational estimators. Journal for Research in Mathematics Education, 13, 183-201.

Riley, R. (1994). Celebrate Goals 2000! Teaching PreK - 8, 24 (8), 12.

Robinson, S. L. (1991). Computer-based instruction in special education. In T. M. Schlecter (Ed.), Problems and promises of computer-based training (pp. 39-60). Norwood, NJ: Ables.

Robinson, S. L. (1994). Classroom technology for the human touch: A mini case study using DISCOURSE™. Curriculum/Technology Quarterly, 4 (1), 1-4.

Robinson, S. L., & Deno, S. L. (1992, November). Computer-enhanced inclusion: A model for computer-based instructional management in mainstream classrooms. Grant proposal submitted to the U. S. Department of Education, Office of Special Education and Rehabilitation Services.

Romberg, T. A. (1992). Mathematics assessment and evaluation: Imperatives for mathematics educators. Albany, NY: State University of New York Press.

Schnaiberg, L. (1994, September 28). '95 budget accord increases education aid slightly. Education Week, p. 23.

Schoen, H. L., Blume, G., & Hoover, H. D. (1990). Outcomes and processes on estimation test items in different formats. Journal for Research in Mathematics Education, 21, 61-73.

Shinn, M. R. (1989). Curriculum-based measurement: Assessing special children. New York: Guilford.

Shriner, J. G., Kim, D., Thurlow, M. L., & Ysseldyke, J. E. (1992). Experts' opinions on national math standards for students with disabilities (Technical Report 3). Minneapolis, MN: National Center on Educational Outcomes , University of Minnesota, College of Education.

Shriner, J. G., Kim, D., Thurlow, M. L., & Ysseldyke, J. E. (1993). Experts' opinions about the appropriateness and feasibility of national math standards (Technical Report 4).

Minneapolis, MN: National Center on Educational Outcomes , University of Minnesota, College of Education.

Sitlington, P. L., Frank, A. R., & Carson, R. (1993). Adult adjustment among high school graduates with mild disabilities. Exceptional Children, 59, 221-233.

Sowder, J. (1992). Estimation and number sense. In D. A. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning (pp. 371-389). New York: Macmillan.

Tindal, G., & Nolet, V. (1995). Curriculum-based measurement in middle and high schools: Critical thinking in content areas. Focus on Exceptional Children, 27 (7), 1-22.

Tindal, G., & Parker, R. (1989a). Assessment of written expression for students in compensatory and special education programs. Journal of Special Education, 23, 169-183.

Tindal, G., & Parker, R. (1989b). Development of written retell as a curriculum-based measure in secondary programs. School Psychology Review, 18, 328-343.

U. S. Department of Education. (1991). America2000: An education strategy. Washington, DC: Author.

U. S. Department of Education. (1993). Goals 2000 Educate America: The National Education Goals. Washington, DC: Author.

Wagner, M. (1990, April). The school programs and school performance of secondary students classified as learning disabled: Findings from the National Longitudinal Transition Study of special education students. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Ysseldyke, J. E., Thurlow, M. L., & Shriner, J. G. (1992). Outcomes are for special educators too. TEACHING Exceptional Children, 25 (1), 36-50.

Zetlin, A. G., & Hosseini, A. (1989). Six postschool case studies of mildly learning handicapped young adults. Exceptional Children, 55, 405-411.

Table 1

Critical Features of Experimental Tasks

Task	Time	Number of Problems	Response Format	Number of Forms Completed per Expt. Session	Sample Problems
Facts Task					
BMOT	1 min.	80	numerical	2	35 + 5 = 3 - 0 = 6 + 8 = 2 * 7 =
Estimation Tasks					
BET	3 mins.	40	single letter (A, B, C)	1	(A) 921 - 480 = A. 2 B. 48 C. 441 (B) Each month I earn $37. How much will I earn in 6 months? A. $3 B. $76 C. $222
MET-A[a]	3 mins.	40	single letter (A, B, C)	2	(A) 7.9 * 9 = A. 16.6 B. 71.1 C. 711 (B) The jacket costs $95. It is on sale for 30% off. About how much will you save? A. $28 B. $125 C. $285
MET-B[a]	3 mins.	40	single letter (A, B, C)	2	(A) 7.9 * 9 is about A. .7 B. 7 C. 70 (B) The jacket costs $95. It is on sale for 30% off. About how much will you save? A. $3 B. $30 C. $300

[a]Note that MET version (A or B) was randomly assigned by class. Students completed either two forms of MET-A or two forms of MET-B.

Table 2

Means and Standard Deviations for General Outcome Measures and Teacher Ratings

	Full Sample			Grade 6			Grade 7			Grade 8		
	N	X	SD	N	X	SD	N	X	SD	N	X	SD
General Outcome Measures												
BMOT	94	11.84	6.03	8	11.19	5.39	48	11.22	5.64	37	12.84	6.71
BET	95	19.86	7.30	8	16.81	8.00	48	18.76	6.94	38	22.04	7.26
MET-A	53	8.66	3.98	4	6.63	5.02	25	8.74	3.66	24	8.92	4.21
MET-B	42	8.50	3.92	4	7.63	2.66	23	7.41	3.65	14	10.57	4.16
Teacher Ratings												
Proficiency	98	3.68	0.86	9	3.44	1.01	49	3.84	0.66	40	3.55	1.01
Communication	98	3.68	0.86	9	3.44	1.01	49	3.78	0.71	40	3.63	0.98
Confidence	98	3.45	0.96	9	3.22	0.97	49	3.63	0.83	40	3.27	1.09
Problem-Solving	98	3.36	0.94	9	3.22	1.09	49	3.49	0.92	40	3.23	0.95
Reasoning	98	3.81	0.86	9	3.78	1.20	49	3.90	0.77	40	3.70	0.88
Value	98	3.34	1.10	9	3.00	1.32	49	3.51	1.04	40	3.20	1.11

Table 3

Correlations Between General Outcome Measures and Teacher Ratings

General Outcome Measure	Teacher Rating					
	Proficiency	Communication	Confidence	Problem-Solving	Reasoning	Valuing Mathematics
BMOT	.52**	.42**	.33**	.54**	.49**	.16
BET	.49**	.45**	.29**	.54**	.43**	.15
MET-A	.39**	.38**	.31**	.42**	.40**	.22
MET-B	.51**	.45**	.19	.50**	.44**	-.12

**$p < .01$.

Table 4

Relative Contributions of General Outcome Measures to the Prediction of Teacher Ratings

Dependent Variable	Independent Variable(s) [a]	Cum. R^2	F Change	p-Value
Teacher Rating				
Proficiency	BMOT	.42	28.25	.000
	BET	--	--	--
	MET-B	--	--	--
Communication	BET	.36	22.34	.000
	BMOT	--	--	--
	MET-B	--	--	--
Confidence	BMOT	.14	6.23	.017
	BET	--	--	--
	MET-B	--	--	--
Problem-Solving	BET	.50	33.11	.000
	BMOT	.53	5.14	.025
	MET-B	--	--	--
Reasoning	BMOT	.43	29.90	.000
	BET	--	--	--
	MET-B	--	--	--

[a] Variables listed in the order in which they entered into the analyses.

Table 5

Relative Effects of General Outcome Measures on Teacher Ratings of Proficiency

Dependent Variable	Independent Variable(s)[a]	Cum. R^2	F Change	p-Value
Teacher Rating				
Proficiency	BMOT	.42	28.25	.000
	BET	--	--	--
	MET-B	--	--	--
Proficiency	BET	.36	21.96	.000
	BMOT	.45	6.59	.014
	MET-B	--	--	--
Proficiency	MET-B	.26	13.39	.001
	BMOT	.43	11.96	.001
	BET	--	--	--

[a] Variables listed in the order in which they were entered into the analyses.

Table 6

Relative Effects of General Outcome Measures on Teacher Ratings of Communication

Dependent Variable	Independent Variable(s) [a]	Cum. R^2	F Change	p-Value
Teacher Rating				
Communication	BMOT	.31	17.18	.000
	BET	.39	5.34	.026
	MET-B	--	--	--
Communication	BET	.36	22.34	.000
	BMOT	--	--	--
	MET-B	--	--	--
Communication	MET-B	.20	10.04	.003
	BET	.37	9.88	.003
	BMOT	--	--	--

[a] Variables listed in the order in which they were entered into the analyses.

Table 7

Relative Effects of General Outcome Measures on Teacher Ratings of Problem-Solving

Dependent Variable	Independent Variable(s) [a]	Cum. R^2	F Change	p-Value
Teacher Rating				
Problem-Solving	BMOT	.45	32.11	.000
	BET	.53	6.02	.019
	MET-B	--	--	--
Problem-Solving	BET	.46	33.11	.000
	BMOT	.53	5.41	.025
	MET-B	--	--	--
Problem-Solving	MET-B	.25	13.18	.001
	BET	.47	15.23	.000
	BMOT	.54	5.93	.020

[a] Variables listed in the order in which they were entered into the analyses.

Table 8

Relative Effects of General Outcome Measures on Teacher Ratings of Reasoning

Dependent Variable	Independent Variable(s) [a]	Cum. R^2	F Change	p-Value
Teacher Rating				
Reasoning	BMOT	.43	29.90	.000
	BET	--	--	--
	MET-B	--	--	--
Reasoning	BET	.36	22.10	.000
	BMOT	.47	7.35	.010
	MET-B	--	--	--
Reasoning	MET-B	.20	9.57	.003
	BMOT	.43	15.98	.000
	BET	--	--	--

[a] Variables listed in the order in which they were entered into the analyses.